Abstracts for EIA's Fall 2007 Meeting with the ASA Committee on Energy Statistics

1. Developing Key Energy Indicators, Janice Lent, SMG, and Joe Conklin, OOG

In June of 2007, an inter-office team of EIA analysts and statisticians was chartered to identify a set of approximately 12 statistics or indicators to be designated Key Energy Indicators. The selected set of indicators will appear together, possibly in a prominent location, on the EIA website for the purposes of

a. providing a quick, easily accessible *overview*—the "big picture"—of the current US energy situation;
b. highlighting, for the media and the interested public, important *changes* in the energy situation; and
c. increasing the accessibility of some of EIA's most broadly relevant products.

1) The key indicators will cover a wide range of energy topics, such as consumption, production, source diversity, use of renewable fuels, and environmental impacts. In this paper, we describe the inter-office team's progress to date in identifying indicators, discuss some related issues the team has considered, and solicit feedback on the team's preliminary list of key energy indicators.

2. Solar Contributions to the Grid, Fred Mayes, CNEAF

Although solar energy produces less than 1% of total U.S. energy consumption, it has received considerable attention over the past 3 decades. The DOE research budget has consistently supported photovoltaic energy research at a high level, while solar energy's clean footprint and theoretical ability to eliminate the electricity grid have made it highly popular among clean energy advocates.

Historically, solar energy applications have tended to be numerous, small, and disperse. Additionally, they require no Federal environmental approvals for construction or operation. These two factors make it extremely difficult and/or feasible for EIA to discover and survey these facilities. EIA also surveys the manufacturers of solar and photovoltaic equipment and uses that information to estimate non-central station solar electricity output--there are few central station solar plants currently. However, several "large" (50 MW) central station solar thermal and PV generating plants have begun operation since 2006 or plan to begin operating in the next 3 years. This raises a major issue of how EIA best uses its electricity and solar equipment survey data to produce comprehensive but non-overlapping estimates of solar electricity in the various sectors of the economy."

3. **Economic Modeling of Natural Gas Processing Plants: Fractionation,** Kara Norman, OAIF

Through the "fractionation" process natural gas processing plants can remove propane, ethane, and butane from the natural gas stream, all of which are marketable. At which point would the owner of that natural gas decide to extract more propane from the natural gas they have purchased? When would the owner opt to leave propane *in* the natural gas stream?

For example if the selling price of natural gas is high, it would be in the plant's best interest to leave as much propane, ethane, and butane *in* the natural gas stream as possible (provided the plant still meets any regulations on the quality of that natural gas). In that case, the volume of the natural gas the owner sells will be larger, they will not need to spend time or money extracting other gases, and thus their overall profit would increase.

The "fractionation spread" so to speak, between the purchased price of natural gas and the potential price or prices of the final products (natural gas, propane, ethane, butane, etc.) is what we are trying to understand.

4. **Comparison of Several Different Non-Disclosure Limitation Methods,** Ramesh Dandekar and Preston McDowney, SMG

A. Comparison of Several Different Non-Disclosure Methods for Tabular Data Using a "Real Life" Table Structure of Complex Hierarchies and Links, Ramesh A. Dandekar SMG

To allow comparison of various sensitive tabular data protection methods on a consistent basis, the statistical disclosure control/limitation (SDC/SDL) researchers have long used public domain artificial (synthetic) data sets available from http://webpages.ull.es/users/casc/ website. The format used by these data sets, however, fails to convey visualization aspects of inherent complexities associated with various structural details typical of public use tables. The practitioners of tabular data protection methods in federal statistical agencies have some familiarity with commonly used table structures. However, they require some guidance on how to evaluate appropriateness of various sensitive tabular data methods when applied to their own table structure. With that in mind, we use a real life "typical" table structure of moderate hierarchical and linked complexity and populate it with synthetic micro data to evaluate the relative performance of four different tabular data protection methods. The methods selected for the evaluation are: 1) lp-based classical cell suppression 2) lp-based CTA (Dandekar 2001), 3) network flow-based cell suppression as implemented in DiAna, a software product made available to other Federal statistical agencies by the US Census Bureau and 4) a micro data level noise addition method documented in a US Census Bureau research paper. The outcome from the comparative evaluation is available from http://mysite.verizon.net/vze7w8vk/
The classical lp-based cell suppression method used for the evaluation is similar to that used by CONFID at Statistics Canada since the mid-80. The selection of the complementary cell suppression pattern is done by using a cost proportional to the table cell value as an objective function. This results in higher preference for smaller tabular cells as complementary suppression cells.
The controlled tabular adjustments (CTA) a.k.a. synthetic tabular data method used is the one documented in Dandekar (2001) and Dandekar/Cox (2002) . Large size non-sensitive table

cells are targeted for adjustments by using a cost function which is a reciprocal of the table cell value. Such an approach results in relatively small percentage changes in the cell values and therefore, reduces the overall degradation in the accuracy of the statistical information imbedded in table cell values.

The network flow model in the DiAna software uses a minimal cost flow (mcf) based algorithm from the University of Texas to develop a complementary cell suppression pattern. The PC version of the software used for this evaluation targets smaller sized cells to develop a complementary cell suppression pattern.
The micro data level noise addition method as described in the paper http://www.census.gov/srd/papers/pdf/bte9601.pdf is used for this evaluation. Micro data is perturbed by an average of 10% and standard deviation of 0.005 by using a normal distribution.

B. EIA Disclosure Limitation Team's Report, Preston McDowney, SMG, principal author w/comments from the Disclosure Limitation Team

EIA collects a variety of energy data under a promise of confidentiality to its respondents. Program offices apply sensitivity rules to protect tabular data from disclosing identity or attribute information about the respondents. EIA has established a statistical standard regarding nondisclosure. (http://www.eia.doe.gov/smg/Standard.pdf) Standard 2002-22, "Nondisclosure of Company Identifiable Data in Aggregate Cells," contains the procedures and policies to identify sensitive (i.e., *primary*) cell values that need protection. Subsequently, disclosure programs, such as Tau Argus (A free software (a result of a European 5th Framework project) developed by Statistics Netherlands' Computational Aspects of Statistical Confidentiality as part of the Fifth Framework of the European Union) and DiAna (A free disclosure software developed by the U.S. Department of Commerce), are used to identify and suppress values in cells for *secondary* confidentiality, values that otherwise could have been used to derive suppressed sensitive values in the tables. EIA tested both software programs using EIA data and tables. This paper will discuss the testing procedures, evaluation criteria, and test results.

5. **Building An Information Quality Page for EIA's Website,** Jacob Bournazian and Lawrence Stroud, SMG

This two-part presentation discusses a proposal to build an information quality page on EIA's web site which would be a comprehensive location for users on information quality issues. The proposal grows out of a discussion from EIA'S October 2005 meeting with the ASA Committee on Energy Statistics when the committee responded to a presentation on investigating data quality issues by comparing the estimates from different data sources. The committee recommended in 2005 to prepare reports relating to data quality issues and release that information as a web based product. This presentation follows up on some of those recommendations and will cover the conceptual and architectural design of a series of web pages for users to access information on information quality. The web pages are in the early stages of development so committee members will be asked about web navigation and usability issues as well as the content and scope of the coverage for building a centralized location on EIA's web site that focuses on information quality issues.

6. Consumption Issues in Industrial Natural Gas Demand, Kobi Platt, EMEU

Natural gas is an integral production component for many prominent industries in the United States. In 2006, natural gas deliveries to industrial consumers measured 6.75 trillion cubic feet (Tcf), or more than 30 percent of the natural gas consumed for the year.[1] EIA's Manufacturing Energy Consumption Survey (MECS), which is completed every four years, examines the components of industrial natural gas consumption. According to the MECS, total industrial natural gas consumption has fallen roughly 13 percent from 1998 to 2002 (from 7.23 Tcf to 6.30 Tcf).

EIA's short-term industrial natural gas consumption forecast is based on industrial production indices which are provided by Global Insight and are based on the Federal Reserve's Industrial Production Database. The production indices of key natural gas-consuming industries are identified using the MECS. Taking the share of natural gas use in each industry that is identified, EIA is able to generate a natural-gas weighted industrial production index.

Since 2002, the trend of industrial production derived from the Global Insight indices has been increasing. However, over the same period, industrial natural gas consumption in these same industries has been consistently declining. Analyses of three prominent natural gas consuming industries—petroleum and coal products, agricultural chemicals, and primary metals—indicates that the associated production indices may no longer provide an accurate representation of natural gas use in that industry.

This presentation will introduce EIA's methodology for constructing a natural gas-weighted industrial production index with relevant background information on the primary natural gas consuming industries that are included. More importantly, the presentation will raise questions about the changing composition of industrial natural gas use in key U.S. industries. Finally, the presentation will conclude with a round-table discussion centered around ways to improve EIA's short term forecasting of natural gas consumption in the industrial sector.

7. Model-Based Sampling Methodology for the New Form EIA-923, Joel Douglas and Jim Knaub, CNEAF

In conjunction with the Electricity 2008 project, the Electric Power Division (EPD) at EIA devised a new sampling methodology for the new Form EIA-923. This form combines aspects of former Forms EIA-906, EIA-920, EIA-423, EIA-767, and FERC-423. Reporting data includes generation and fuel consumption, stocks, receipts, cost, and quality. The sampling strategy focuses on sample reduction with special attention placed on ensuring adequate monthly coverage of reported data. Sample reduction is deemed especially important for the commercial and industrial sectors due to the respondents' higher difficulty in reporting for those facilities as well as to the questionable data quality of some of the smaller respondents. The basic sampling methodology employed for the new form is a modified cutoff sample based on operating generating capacity. Additions to the sample are included to ensure acceptable

[1] Energy Information Administration. *Natural Gas Monthly: Annual Natural Gas Consumption by End Use*. http://tonto.eia.doe.gov/dnav/ng/ng_cons_sum_dcu_nus_m.htm. April 2007.

regional coverage of generation, consumption, and stocks base. Counts for established estimation groups, and relative standard errors are also analyzed to ensure they meet acceptable thresholds. EPD has endeavored to consider both sampling and nonsampling error in this study.

8. The Quality Control Process for EIA Projections and Data, Paul Holtberg, OIAF and William Watson, CNEAF

A. The Quality Control Process for Developing the Annual Energy Outlook, Paul Holtberg, OIAF

The Energy Information Administration follows a detailed review process prior to release of its major products. The most extensive review is typically reserved for those EIA products that receive the most visibility. This presentation will discuss the process used for quality control/review for the Annual Energy Outlook (AEO). The review and quality control process for the AEO begins with reassessment of new historical data to update model relationships and to develop the assumptions which are typically external to the model dynamics such as a world oil prices and other non-U.S. market behavior. Such external assumptions are vetted within OIAF and then the rest of EIA through working groups and a Delphi process. Once the major external model assumptions are developed, model runs are examined carefully within OIAF at all levels of staff and management and discussed at weekly or bi-weekly "run review meetings." When the major cases of the AEO are considered satisfactory by OIAF management and staff, they are once again vetted through the rest of EIA, including the Administrator and Deputy Administrator, the EIA Offices, and management and analysts in other Offices. The analysis and documentation of the runs and the resulting writeups follow a similar pattern of review, including both internal Office review and wider EIA review. After the AEO is released, feedback from the various stakeholders are solicited to provide input to the next production cycle.

B. What Processes, Reviews and Procedures Does CNEAF Follow to Assure Quality and Release Data? William Watson, CNEAF

CNEAF conducts national data surveys to collect data for electric, nuclear, coal, and alternate fuels markets. Within CNEAF, the Coal Team conducts two annual surveys: (1) form EIA-7A sent to coal mines to collect data on production, reserves, and coal sale revenues; and (2) form EIA-6A sent to coal distributors to collect data on coal distributed to major consumers by transport mode, origin, and destination States. In addition, the Coal Team conducts two quarterly surveys: (1) form EIA-3 sent to manufacturing plants to collect data on coal receipts, consumption, stocks, and coal cost and quality; and (2) form EIA-5 sent to coke plants to collect data similar to that collected on the EIA-3 form plus data on coal by-products produced in coke plants. This presentation reviews various processes that CNEAF has put in place to assure full frame coverage and data quality for its coal surveys. The presentation evaluates frame processes, survey form design, data edit flags, resolution of edit flags, and aggregate-frame editing as processes operating to assure accuracy of the surveyed data. Strengths as well as weaknesses are discussed. Enhancements are proposed to improve the QA processes and recommendations from the Committee are sought in terms of priority and additional steps the Committee may identify.

.

9. **Evaluating Model Quality: Features of the Next Generation of Model Quality Audits,** Nancy J. Kirkendall, SMG

During the original development of EIA's National Energy Modeling System in the early 1990's, the predecessor of the Statistics and Methods Group provided a Model Quality Audit (MQA) program that evaluated modules as they were completed. The purpose of the MQA program was to "evaluate the soundness of the model's mathematical/statistical and economic foundations." The steps in the old audits that are considered for inclusion in the new program include:

1. Assess the model methodology and associated economic, mathematical, and/or statistical foundations of the model through Independent Expert Reviews (IERs) within one year from the time a significant set of modeling changes are made.

2. Evaluate the mathematical and statistical properties of the model. This might include sensitivity analysis, uncertainty analysis, evaluation of model specification errors and model bias, partitioning error into categories such as error due to assumptions in conditioning variables, and those due to changes in laws, and evaluation of regression diagnostics.

3. Verify the documentation and implementation of the model: (a) Is the model documented according to EIA standards? (b) Does the archived model operate as intended? (d) Is the implemented model consistent with the documentation? (c) Is the model archived according to EIA standards?

Now that EIA is launching its NEMS redesign, it is time to determine what a new MQA program might look like, and what role the Statistics and Methods Group should play. In 1990, the predecessor to SMG conducted Independent Expert Reviews during model development to inform and improve model design, and focused on evaluation of models after they had been developed and implemented by OIAF. In 2008, we could also consider helping in the early stages of NEMS redesign by organizing workshops or focus group discussions to obtain external advice.

In this presentation, Dr. Kirkendall will describe EIA's current ideas about the MQA process, and will solicit advice from the Committee concerning activities that should be included.

10. EIA Forecast Evaluation Using Regression Analysis, George Lady and Andrew Buck, Consultants to SMG

The overall project goal is to assess the sensitivity of NEMS projections to the values assumed for important, related explanatory variables. The sensitivities are to be estimated via regression

analysis of NEMS solutions specifically configured to identify the impacts of changing the assumed values. This report presents regression results for the demand for selected fuels for each of the residential, commercial, industrial, and transportation sectors. Analysis of the NEMS solutions for which individual fuel prices were changed, one at a time, revealed that cross-price sensitivities were extremely small. Accordingly, the regression specifications utilized were austere: estimating the demand for each fuel as a function of the fuel's own price and one or more related activity drivers, e.g., the residential consumption of electricity as a function of the price of electricity and the total number of residential households. An inter-temporal profile for the sensitivities found was provided by including the lagged endogenous variable on the right-hand-side of the regression specification.

Since the number of explanatory variables was small, the experimental design of the solution sets used to provide data for the regression analysis could be configured using the Latin hypercube criterion for each explanatory variable being assigned a base, high, and low value. The fit of the regression results was quite good. The implied elasticities, and their inter-temporal variation compared closely to the values found in a nominal evaluation of the data.